JIARUI LUO

Tel: +86 15959013976| Email: Jiarui.Luo@outlook.com | Email: luojr2019@mail.sustech.edu.cn

GitHub: https://github.com/SpaceIshtar | Homepage: https://spaceishtar.github.io/

EDUCATION

Southern University of Science and Technology

Bachelor in Computer Science and Engineering

- GPA: 3.89/4.00, Rank: 3/157 •
- Key courses: Probability and Statistics, Data Structure and Algorithm, Database Principles, Computer Architecture, Information Theory, Data Mining, Machine Learning

RESEARCH INTEREST

System and algorithms for large-scale data.

PUBLICATION

R. Guo, X. Luan, L. Xiang, X. Yan, X. Yi, J. Luo, Q. Cheng, W. Xu, J. Luo, F. Liu, et al., Manu: A Cloud Native Vector Database Management System, The International Conference on Very Large Databases 2022.

RESEARCH EXPERIENCES

Research Topic: Learning Methods for Database Tuning

Role: Research Assistant, Supervisor: Prof. Guoliang Li in Tsinghua Database Group

- Database has hundreds of knobs that significantly influence the performance of a database. In recent years, a number of learning based methods have been proposed to tune the knobs automatically.
- o Investigated and implemented database tuning methods, including OtterTune, CDBTune, and QTune, etc.
- o Explore the possibility of transferring tuning models from one tuning environment to another, in order to reduce the tuning time for database users.

SIGMOD'22 Programming Contest on Entity Blocking, Finalist

Role: Leader, Supervisor: Prof. Bo Tang and Prof. Xiao Yan in Database Group

- Entity blocking, a step for entity matching task, is used to quickly filter out the non-matches. The contest prepared million-scale datasets from e-commercial websites, and participants were required to design a system that generates small candidate sets of pairs with high recall. Our solution ended up with a finalist award.
 - o Led a team of three students, proposed to use pre-trained neural network model to transform entity description sentences into representative vectors.
 - o Adopted a HNSW index, increased graph connectivity in HNSW to improve performance and limited the size of priority queue in the searching process to reduce search time, as a result the similar vectors were matched quickly with high probability.

Research Topic: Graph Summarization

Role: Leader, Supervisor: Prof. Bo Tang and Prof. Xiao Yan in Database Group

- Given a graph, graph summarization represents it as a summary graph and edge corrections, with the goal of minimizing the overall size of the two parts. As a result, graph summarization could be used for compression and pattern mining. I proposed a new solution that improves the compression rate and execution time.
 - Investigated graph summarization problem, tested related works, i.e., Randomized, SWeG, LDME.
 - Assessed different hashing schemes, including MinHash, DOPH and Asymmetric LSH.

Shenzhen, China Sept.2019-Jun.2023

Mar.2022-May.2022

Oct.2021-Feb.2022

Jun.2022-Current

- Explored node encoding and maximum inner product search to make improvements of compression rate.
- Designed a LSH tree structure after analyzing the problem in detail to solve overfitting problem. Proposed an algorithm based on the LSH tree structure to summarize graph more efficiently.
- Conducted experiments and proofed the superiority of the algorithm on compression rate and elapsed time.

NeurIPS'21 Billion-Scale Vector Search Competition (Track 2), Champion July. 2021-Oct. 2021 Relay Team Member Supervision, Prof. Po. Tang. and Prof. Vigo Van in Database Crown

Role: Team Member, Supervisor: Prof. Bo Tang and Prof. Xiao Yan in Database Group

- Track 2 conducts vector search on 6 billion-scale datasets with 64GB DRAM and 1TB SSD, and compares the recall obtained at 1500 query/s. The key is to optimize data read for the low bandwith and long latency of SSD. Our proposal boosts the recall competition baseline on range search by more than 70%.
 - Investigated, implemented and assessed key vector search methods, including Locality Sensitive Hashing (LSH), Proximity Graph, Vector Quantization, etc.
 - Investigated the characteristics of SSD I/O in detail to make fully use of SSD.
 - Proposed bucket graph solution for billion-scale datasets that clusters vectors into 4KB buckets to align with SSD blocks for efficient read. Built HNSW to navigate among the bucket centers in DRAM.
 - Implemented scalar quantization for read reduction, HNSW for bucket center search and Python-C++ interfaces that connected our implementation with the competition test environment.
 - Conducted experiments to verify the superiority of our solution, which boosts the recall of competition baseline on range search by more than 70%.
 - Applied our solution to Milvus, a cloud-native vector database system, which has been adopted by many enterprises including Tencent, ebay and NVIDIA.
 - \circ Published a paper that describes our solution on VLDB 2022 conference about the system.

TEACHING EXPERIENCE

Southern University of Science and Technology	Feb.2021-Current
Role: Teaching Assistant	
Machine Learning, Fall 2022 by Professor Qi Hao	
Created and maintained an online judge system	
• Helped international students in the lab and after lectures.	
Discrete Math, Spring 2021 by Professor Adam Ghandar	
Helped develop assignments	
Answered questions after lectures	
Graded assignments and midterm exams	

HONORS & AWARDS

•	Chinese National Scholarship	2022
•	Finalist, SIGMOD'22 Programming Contest on Entity Blocking	2022
•	Honorable Mention, The International Mathematical Contest in Modeling	2022
•	Champion, NeurIPS'21 Billion-Scale Vector Search Competition	2021
•	3rd Prize , University Merit Student Scholarship	2021
•	2 nd Prize, College ACM Contest	2020
•	2 nd Prize, University Merit Student Scholarship	2020

SKILLS & TOOLS

Programming Languages: Java, C/C++, Python, SQL, Matlab, Verilog HDL
English Proficiency: IELTS (7.5), TOEFL iBT (105), GRE General Test (325)
Tools: Jetbrains (IntelliJ IDEA, Clion, PyCharm, DataGrip), VSCode, Jupyter Notebook, Unity, Ubuntu